Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Improving the efficiency and reliability of wearable based mobile eHealth applications

Cesar Garcia-Perez [a], Almudena Diaz-Zayas [a], Alvaro Rios [a], Pedro Merino [a,*], Kostas Katsalis [b], Chia-Yu Chang [b], Shahab Shariat [b], Navid Nikaein [b], Pilar Rodriguez [c], Donal Morris [c]

[a] *Universidad de Málaga, Andalucía Tech, Spain*
[b] *Eurecom, Communications System Department, France*
[c] *RedZinc Services Limited, Ireland*

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
mHealth
IoT
Sensor
Wearable
Mission critical
LTE

ABSTRACT

In this paper we address the support of wearable mHealth applications in LTE and future 5G networks following a holistic approach that spans across the elements of a mobile network. The communication requirements change from one application to another so we propose a measurement methodology to facilitate the selection of the user equipment to fulfil these requirements. We also discuss a new network architecture to support traffic prioritization, RAN programmability, low latency and group communications to over-the-top applications. Our proposal is validated using several realistic experimentation platforms and the results show that mHealth systems can benefit from our approach.

1. Introduction

The trend in mobile networks used to support healthcare services was recognized by the World Health Organization in 2011 [1], where they defined the concept of *mHealth* as *"the communication or consultation between health professionals about patients using the voice, text, data, imaging, or video functions of a mobile device"*. One key finding of this report was that *"The Americas (75%), European (64%) and South-East Asia (62%) regions reported high rates of adoption of mobile telemedicine initiatives, though a large proportion of these initiatives were informal or in the pilot phase"*. Since that report, in just five years, many more opportunities for mHealth have arisen due to the evolution of the Internet of Things (IoT), the deployment of LTE networks and the definition of the requirements for future 5G networks. One emerging application domain is the use of 5G wearables and other sensors technology, continuously reporting data on a patient's state over a mobile network to the hospital or to the command centre in case of disasters.

Wearable mHealth applications span across many different network scenarios, like very dense sensor networks providing real time information of users or real-time streaming video transmissions to support early diagnosis on field. Mobile communications standardization bodies are increasing efforts to provide technology to support these type of applications, which can have diverse requirements. Some discussion about the requirements of mHealth applications is provided in [2] and [3].

To successfully deploy these wearable-based mHealth applications on a massive scale we need to provide specific data transmission mechanisms in networks to ensure aspects such as performance, network efficiency and reliability. For

[*] Corresponding author.
E-mail addresses: garciacesaraugusto@lcc.uma.es (C. Garcia-Perez), pedro@lcc.uma.es (P. Merino).

http://dx.doi.org/10.1016/j.pmcj.2017.06.021

instance, mobile networks are usually optimized to increase the speed of file or video download (from the network to the user), but IoT sensors and wearables require good quality transmission in the uplink. In addition, other requirements are related to low latency communications (time to travel the network from the sensor to the consumer), traffic prioritization schemes (obtaining the right to send data in case of congestion in dense areas) or seamless handover (transparent change from one cell to a new cell). However, such specific requirements, which are also common in other mission critical applications, are still not supported by the commercial mobile networks (see [4]).

This paper presents an approach to overcome some of the challenges posed by mHealth communications, as laid out in the European project Q4Health [5]. To do so we propose techniques to optimize the services both from the wearable and network perspectives. From a wearable point of view we propose a methodology to select the most appropriate User Equipment (UE). Experience has taught us that UE data sheets are not enough to make an optimal decision as they cannot provide a realistic view of the final behaviour of the application. The impact of the device driver, the platform, the traffic profile cannot be ignored when selecting the UE. With this approach we can provide insights into the affect of the execution platform, on the behaviour of the platform with different mobility patterns, the latencies both in the data and control plane and the energy consumption of the full system.

Regarding the network, we have provided an architecture that, in an API network, exposes functionality to third parties. Using this API mHealth applications can request QoS from the network to protect their traffic, ask for low latency or group communications and exploit the RAN programmability to better adapt the network scheduling to their applications. Low latency services are provided by Mobile Edge Computing (MEC) architectures with different approaches, exploiting the data plane as in [6,7] or exploiting the control plane information. With the same architecture we can provide more efficient group communications by bypassing the core network. In order to implement traffic priorities two different approaches are provided, one exploiting the already existing functionality in LTE networks and another aimed at 5G networks. In the latter we work in the MAC scheduler using an SDN-based approach with a controller in charge of applying policies to the schedulers. All these functionalities can be requested/configured by third parties, with the proper technical and business agreements with the network operator, and are exposed in an API so that they can be accessed programmatically.

The whole approach is validated with a set of experiments over two experimental platforms PerformNetworks [8,9] and OperAirInterface [10].

The rest of the paper is organized as follows: Section 2 provides a state of the art of the different topics explored in this paper. The overall optimization approach and the validation methodology are described in Section 3. Section 4 provides an overview of the optimization procedures to select the best UE, which covers the behaviour on the cell edge, the data plane latency, the power consumption and the throughput. Section 5 is devoted to the proposed network architecture to improve mHealth applications, providing details on the QoS enforcement, the latency reduction techniques, the optimized group communications and the RAN programmability to support the scheduling techniques employed in the platform as well as some results on downlink optimization. Section 6 analyses the results and discusses their strong and weak points. Finally Section 7 details conclusions and future work.

2. Related work

In this section we first discuss existing work on IoT and critical services in mobile networks; then we give an overview of scheduling strategies in the base stations. The state of the art in low latency solutions is also analysed.

2.1. Cellular networks to support critical services and IoT

Moving towards integrated 5G communications and driven by the need to satisfy extreme requirements for critical applications, like healthcare, the research community is reconsidering the role of commercial cellular networks. In [4] a detailed study is provided by the European Union, analysing the use of broadband communications for mission-critical and high-speed applications. The interested reader is directed to FirstNet [11] for related activities in the US. In [12] broadcast communications for public safety LTE networks are described, using a number of different options for multicast resource allocation. The proposed schemes are used to secure optimal radio resource efficiency using a mix of synchronous and asynchronous multicast traffic, and provide public safety services with minimal interruptions due to mobility. This is also discussed in [13], which focuses on the use of the functionality for mission critical communications. The effects of mobility on critical communications are also studied in [14], while in [15] Maskey et al. provide a comprehensive latency analysis for M2M applications in LTE networks.

2.2. LTE scheduling strategies

In LTE networks the most important factor that affects end user performance is the way Resource Block (RB) allocation is performed at the MAC layer. Some general discussions on LTE scheduling strategies can be found in [16]. Algorithms are categorized as (a) channel independent; (b) channel sensitive, where a distinction exists when the objective is (b1) to maximize the QoS requirements of each UE (QoS aware scheduling) or (b2) to be fair among UEs (QoS unaware scheduling). Other categories include delay-based algorithms and power-based algorithms. A survey for the uplink case and SC-FDMA is presented in [17], which is of particular interest for the video uplink transmissions that are proposed in this paper. In [18]

Please cite this article in press as: C. Garcia-Perez, et al., Improving the efficiency and reliability of wearable based mobile eHealth applications, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.06.021.

a three-step iterative scheduling is proposed where scheduling is done on per-user-per-service basis. The MAC scheduler analyses the parameters associated with all the logical channels (service radio bearers) for each user and applies a scheduling algorithm to the packets of logical channels individually rather than just at the user level. The authors of [19] proposed a similar approach in two steps, validated with a simulator. In [20] a scheduling algorithm based in QoS and DRX cycles is proposed for VoIP applications, and the approach is validated with a system level simulator. In [21] the problem of Resource Blocks allocation is again researched for the downlink, where the authors present the case of single user optimization with the optimal MCS allocation and multi-user throughput optimal scheduling.

2.3. Optimization of the network to support critical services and e-Health

Related work for optimizing mobile networks that support eHealth services and also involve IoT devices, considers the use of Software Defined Network (SDN) technologies to configure the behaviour of the components in the mobile network. Nguyen et al. provide a comprehensive survey of SDN for these networks in [22]. Non-standard approaches usually focus on the data plane such as the dynamic allocation of resources to tunnels in the network [23]. Programmable SDN underlay with a focus on Healthcare applications is also researched in [24] where a Health IoT solution is proposed.

The authors in [25] advocate the use of LTE gateways for mobile health environments. This is the most common approach in the literature with regard to pervasive computing systems [26]. Note however that with the coming of new IoT standards (i.e., NB-IoT [27]), sensors could be directly connected to the network thanks to miniaturization and power consumption optimization. IoT gateway systems could exploit new standard ways to optimize the data path in mobile networks, and to consider new cloud-based designs (like the MEC [28]), which have been applied to optimized video streaming services in [29]. Another concept similar to MEC is FOG computing, which has been analysed for Internet of Things applications in [30].

3. Optimization of mHealth applications

This section provides an overview of our approach, which is based on the optimization of the UE as well as network evolution. Our approach is driven by a use case, a wearable video platform for emergency services, which is described further on. Finally we describe the validation methodology that has been followed.

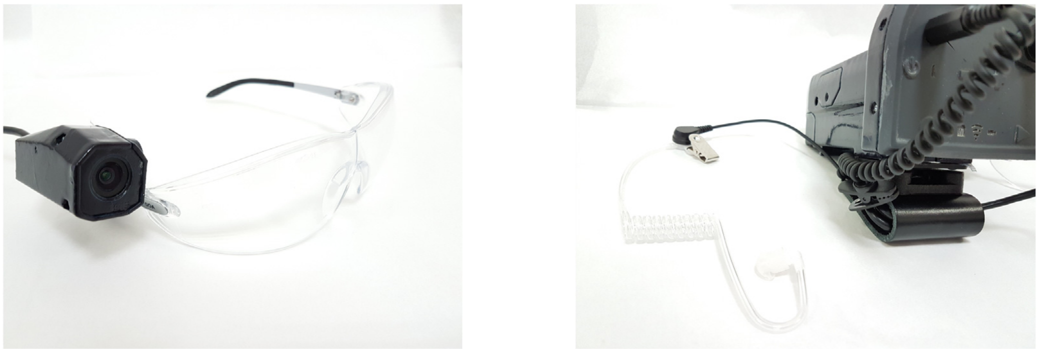
3.1. Approach

In order to improve the mHealth communications we employ a holistic approach and try to optimize both the network and the components of the integrated communication system that supports the mHealth applications.

One important aspect for the mHealth application is the selection of an appropriate UE. Usually over-the-top solutions integrate their own communication systems (as opposed to those running over consumer mobile devices). This is particularly useful for wearable systems, which are normally integrated with different devices to optimize the overall performance. The selection of the UE has to be done taking into the account many different performance indicators, as well as considering other factors such as cost or size. To provide realistic data about the terminals we define a set of measurements aiming to support an optimal selection, which are:

- Maximum Output Power and Receiver Sensitivity: these are used to characterize the behaviour of the UE on the cell edge. The DUT maximum output power influences the performance of the uplink connection for the mobile users on the cell edge and other scenarios with poor coverage; thus for applications with high mobility it is an important parameter. Additionally the receiver sensitivity affects the performance of the UE for the downlink data receptions close to the cell edge, as the UE will receive very low power.
- Power Consumption: is a key factor in eHealth and IoT scenarios where autonomy is very important. We investigate the instantaneous power consumption and also estimate of the battery life by forcing the application to work in different network conditions.
- Throughput: in order to provide a realistic characterization we estimate the throughput under different mobility patterns, also obtaining the jitter and packet loss.
- Control Plane Latency: the control plane procedures affect the overall latency of the data plane. We provide measurements for the attach procedure, the service request and the dedicated bearer establishment procedure. The attach procedure comprises all the signalling required when the UE has to register in the network. It performs the identification and verification of the UE SIM, the establishment of security, the allocation of resources, and the capabilities negotiation. The service request is a procedure that is triggered when the UE is in energy saving mode (named idle mode) and needs to transmit data. This procedure is particularly relevant for devices with long periods of silence, which go to idle more easily. Finally the dedicated bearer establishment procedures are also analysed as they are relevant for applications requiring traffic prioritization.
- Data Plane Latency: we provide some measurements obtained with applications that analyse the Round Trip Time (RTT) of ICMP packets in each of the segments of the network. Additionally we extract information of the time taken by each of the segments of the network.

Please cite this article in press as: C. Garcia-Perez, et al., Improving the efficiency and reliability of wearable based mobile eHealth applications, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.06.021.



(a) End user glasses.

(b) Belt-mounted wearable.

Fig. 1. BlueEye platform detail.

These measurements for the Device Under Test (DUT) can be acquired using it in the final platform base and in order to facilitate the device selection by highlighting its behaviour in different situations. The detailed methodology as well as results of a comparison of two different devices are provided in Section 4.

Regarding the network optimization we propose an enhanced LTE architecture, geared towards 5G evolution, which provides functionality to support prioritization, reliability, low latency and group communications. All the functionality is exposed in an API that can be called by third party applications in order to trigger it. This improves the flexibility of the network and eases the deployment of critical mHealth services, simplifying the interaction with the operator programmatically.

To support QoS enforcement on LTE networks, the API exploits the PCRF Rx interface [31], which was traditionally employed by the IP Multimedia Subsystem, in order to trigger QoS demands to the network. The QoS demands on the Rx interface will trigger dedicated bearer establishment in the network, so some figures on the use of dedicated bearers with commercial base stations are also provided.

Low latency services are provided by MEC architectures, two principal types of solutions are provided, one based on the data plane analysis and another on the control plane. The data plane solutions analyse the GTP traffic to decide which packets have to be addressed to the fog, and which can provide sufficient performance but requires downlink data to work. To avoid this requirement the control plane analysis can be placed in the architecture. The same MEC architecture can be exploited to support group communications, providing the distribution of the data for the group in the MEC servers, avoiding part of the traffic towards the EPC. This scenario is useful for sensor network communications and to support multicast communications of small groups.

The RAN programmability is supported by exploiting the FlexRAN, which is a programmable framework. Through the separation of the RAN control and data plane and aided by virtualized control functions and control delegation features, FlexRAN provides a flexible control plane that is designed with support for real-time control applications, programmability and flexibility for different degrees of coordination among RAN infrastructure elements. The FlexRAN solution is also extended to support RAN slicing capabilities. Using the enhanced FlexRAN API, a dedicated network slice for the mHealth application can operate in parallel with other network slices, with guarantees on performance and isolation.

3.2. Motivating use case

The validation of the integrated solution is done via a specific mHealth use case demonstration. In the integrated system, the network and system optimization procedures developed are demonstrated for an eHealth real-time video wearable platform named BlueEye. The platform consists of mounted safety glasses (see Fig. 1) that provide high definition video to paramedics and emergency services. The video is sent over an LTE connection and connected to a backend that provides a measurement framework, video distribution services and the VELOX system, which allows access to all third party functionality such as end-to-end orchestration.

The BlueEye platform provides the driving use case for the technological innovations of the project. Paramedics will use the wearable video platform to send video of a patient to a Hospital, where specialists can provide an early diagnosis. The video will be transmitted on site, in both outdoor and indoor scenarios, and from pedestrian fading profiles to vehicular ones when the paramedic moves from the field to the ambulance. Additionally we have foreseen an emergency scenario where a hospital campaign is deployed and the video from the emergency services is transmitted both to the field and fixed hospitals.

Please cite this article in press as: C. Garcia-Perez, et al., Improving the efficiency and reliability of wearable based mobile eHealth applications, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.06.021.

Table 1
Comparison between DUT A and DUT B.

	DUT A	DUT B
Sensitivity (dBm)	−123.1	−126.6
Maximum output power (dBm)	21.6	22.5
Power consumption (mW)	0.5303	0.5408

3.3. Methodology

To validate our approach we have used two experimental platforms PerformNetworks [8,9] and OpenAirInterface [10], both featuring realistic LTE and 5G equipment. The platforms have been used to perform different experiments to provide a characterization of the methodology and the network improvements proposed.

The UE selection methodology has been performed using Keysight UXM and T2010 units, which are pieces conformance testing equipment, combined with Spirent channel emulators. The conformance testing units can emulate a base station but with full control of the parameters of the LTE-A protocol stack. The units support downlink channel emulation and to introduce uplink fading and noise we use the channel emulator. This equipment has been combined with a LTE core network emulator that provides functionality for all the core network elements depicted in Section 5.1.

To analyse the effect of some of the expected improvements of future 5G networks several prototypes are currently being developed and integrated into the testbed and made public for external applications. Furthermore, several proof of concepts have been carried out employing commercial base stations and packet core networks.

Note that the open-source OpenAirInterface (OAI) system can also be used to generate low level measurements and is able to provide the CQI, RSRP and RSRQ measurements calculated by the UE. Besides these measurements the energy consumption from the terminals can also be highly relevant in mobile eHealth applications. A power analyser is used to supply energy to the UEs so samples from the current and voltage can be taken to estimate the instantaneous power consumption. For the RAN programmability scheduling OAI eNB running over USRP B200mini software dened radio (SDR) cards and is combined with a third party EPC.

Another important aspect of the UE is its behaviour under different signalling procedures, which includes what signalling procedures are supported and how long it takes to finish them. Several tools for analysing the signalling messages from the EPC have been implemented and used to characterize this behaviour. The tools are able to capture and dissect the messages of the S1AP protocol, which provides the signalling service between the base station and the EPC in the S1 interface.

Finally, besides the application's specific KPIs, several common application-specific KPIs are also measured. These include the data plane latency, which includes the latency introduced by the radios and the EPC components, throughput and jitter .

4. Selecting the Optimal UE

In this section we provide details about the UE selection methodology previously defined in Section 3.1, which can be used to ease the selection of the most appropriate UE for a given mHealth application. Two different devices have been evaluated following the proposed approach, hereinafter referred to as Device Under Test (DUT) A and DUT B. Specifically the comparative evaluation aims to characterize the behaviour on the cell edge, the power consumption, the latency on the data plane and the throughput. The importance of these characteristics can vary depending on the specific mHealth application.

4.1. Characterizing the behaviour on the cell edge

The behaviour on the cell edge is very important for wearable applications, which will normally be subject to high mobility. The DUT maximum output power and the receiver sensitivity have been used to characterize the uplink and downlink transceiver performance respectively.

4.1.1. UE maximum output power

LTE transmitter measurements are defined in the specifications [32] and [31]. The UE maximum output power test measures the power transmitted by the DUT in the channel bandwidth when it is instructed to raise its power constantly until limited by its power class. According to the specifications, DUT maximum output power must be measured using QPSK modulation and partial resource block (RB) allocation. This will be the most frequent situation on the cell edge as bad channel conditions will limit the uplink scheduling grant and the UE will reduce its modulation index due to these bad channel measurements. This is done to meet the need for a robust modulation when the uplink signal cannot achieve a suitable level at the base station because the power limits have been reached. For 20 MHz bandwidth the specification indicates that the maximum number of allocated RB is 18. In these conditions the maximum output power obtained for the devices under test is provided in Table 1.

Table 2
IP performance measurements.

	Ideal DUT A	Ideal DUT B	Pedest. DUT A	Pedest. DUT B	Vehic. DUT A	Vehic. DUT B
Throughput (kbps)	49,715	49,716	31,088	37,709	9967	13,146
Jitter (ms)	0.42	0.47	3.75	2.05	3.78	3.54
Packet loss (%)	0	0	8.39	6.64	36.8	26.46

4.1.2. Receiver sensitivity

The worst case for downlink reception on the UE side is when the device receives a very low signal power, which is typically called a sensitivity level test. When operating at very low power, the mobile device will mainly be affected by its own noise in addition to external interferences. In order to minimize the impact of external interferences a shielding box, which provides strong isolation from external impairments, has been used. This way the mobile device receiver can be better characterized. The following custom test sequence has been defined:

- Set the UE to its maximum power increasing transmit power control.
- Set the downlink power to −85 dBm/15 kHz (default in the testing standards).
- Enable Physical Downlink Shared Channel (PDSCH) transmission with Modulation and Coding Scheme Index (IMCS) 5. This causes QPSK to be used, which is consistent with the need for a robust modulation at the cell edge.
- All the subframes except 0 and 5 transmit in the downlink with this format, as stated in [33].
- Measure the average throughput during the time required to achieve statistical confidence as per the conformance specification.
- In 1 dB steps, reduce the power and repeat the throughput measurement while the throughput exceeds 95% of the maximum rate.
- When the throughput falls below that threshold, the power is increased 1 dB and the same process is repeated with steps of 0.1 dB.
- In this second finer loop, the last power value where the throughput exceeds the 95% threshold (BLER below 5%) is stored as the sensitivity value.

Following this approach we have obtained a sensitivity of −123.1 dBm for DUB A and −126.6 dBm for DUT B.

4.2. Power consumption

Another important characteristic that is part of our testing methodology is related to power consumption. Power consumption testing is very important and directly impacts the battery life of the UE and it is a factor that usually varies between different vendors.

In order to characterize the power consumption for each device, a set of experiments were carried out and the results averaged over 5 test iterations. The power consumption samples were produced every 100 ms. Each test execution spans over 120 s, where constant rate UDP traffic was generated at 50 Mbps in the uplink to ensure that the devices are operating at their full capacity.

The test conditions have been selected to replicate a bad coverage scenario with high propagation losses. In such a scenario, the devices under test are expected to transmit at their maximum available power to compensate the channel propagation losses they estimate. In this case, very similar results for both DUTs have been obtained, with minor differences in terms of average power consumption. The results are provided in Table 1.

4.3. Throughput performance

In order to assess the IP data performance with both mobile devices under impaired conditions, multiple experiments seeking to reproduce scenarios with high propagation losses have been conducted. The conditions of the experiments are intended to reproduce a worst case scenario where the mHealth service could be used in locations with poor coverage conditions, where the communication integrity could be at risk and following different mobility patterns (static, pedestrian and vehicular). When fading due to mobility is present, Additive White Gaussian Noise (AWGN) has been added to force an statistical mean SNR of 20dB when the devices transmit at 23 dBm.

In accordance to test recommendations for application data testing, we have repeated each experiment 5 times. Table 2 provides a summary of the results obtained during these experiments while Fig. 2 depicts the average throughput represented in intervals of one millisecond.

According to the capability of the devices under test, they are capable of generating up to 51,024 Kbps at the LTE physical layer, which is consistent with the results obtained at application level considering the headers introduced at the different protocol layers. It can be observed that under static scenarios (scenarios without fading) the maximum uplink throughput can be achieved for the whole test period.

As intuitively expected, the results obtained in the pedestrian scenario are better than in the vehicular case, as the former's propagation profile reproduces low speed conditions, and a lower number of multipath reflections. For the vehicular scenario

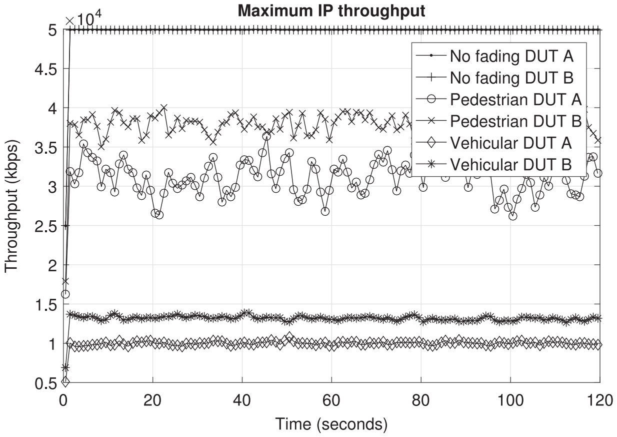


Fig. 2. IP traffic measurements.

Table 3
Summary of the data plane analysis results.

	Median RTT (ms)	MAD RTT (ms)
DUT A	88,71	4,936
DUT B	92,714	5,004
DUT A (fading)	89,942	5,464
DUT B (fading)	93,733	5,061

the throughput degrades by a considerable percentage in respect the theoretical maximum due to the high Block Error Rate (BLER) obtained in the uplink. The mean BLER in such a scenario is around 70%, while the maximum number of HARQ (Hybrid ARQ) retransmissions has been configured to 4. This is a working point used to measure the available throughput in extremely impaired radio conditions.

In the pedestrian scenario, the mean BLER is around 20%, RF conditions are still poor, which causes a significant degradation of the throughput. Usually, BLER needs to be better than 10% in order to consider that radio propagation conditions as favourable. Additionally, the lower profile speed causes instantaneous transmission power variations to be slower over the time, causing a more noticeable ripple in the represented throughput for the pedestrian scenario.

Interestingly, despite having a lower throughput in both devices, the vehicular scenario seems to present a lower ripple. This is caused by the averaging interval, because the vehicular profile fluctuates at a higher speed but maintains the same statistical mean in the long term. It can be observed that the generated throughput is slightly but noticeably and consistently higher in the second device under test in all propagation conditions except in the ideal conditions where both devices reach the maximum.

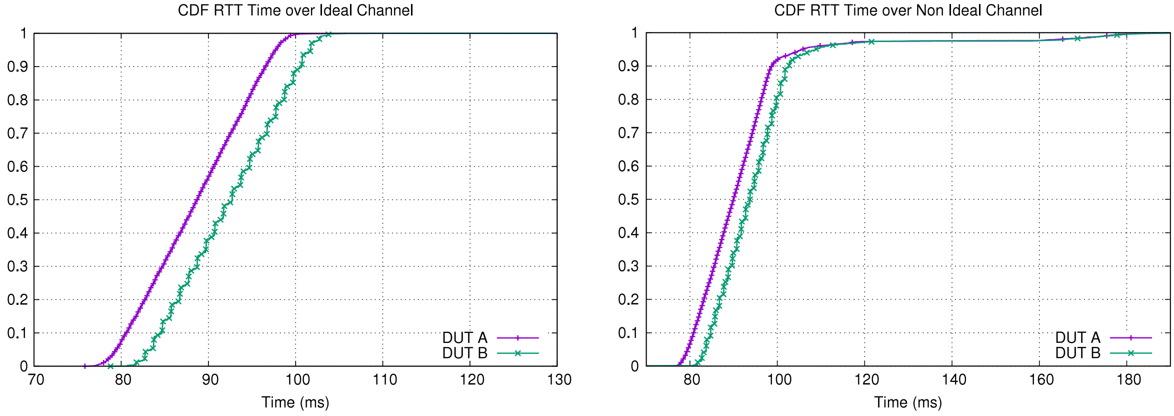
This is strongly aligned with the results obtained in Table 1, where the second device has been measured to transmit approximately 1 dB higher when tested at maximum power.

4.4. Latency of the data plane

Another important aspect of communications is the latency introduced by the components of the network. Several solutions to reduce it are described in Section 5.3. Here a baseline of the delay introduced by the radio access of each DUT is provided.

The results have been generated connecting the DUT to conformance testing equipment modified so it could be connected to an EPC. In the EPC delay impairments are introduced in the S1-U (data plane interface with the eNB) and SGi (reference point that connect the EPC with external networks). To characterize the KPI a tool has been developed. The tool analyses PCAP traces of the UE, the S1-U interface and the SGi interface and provides an estimation of the RTT based on ICMP traffic. The UE is configured to generate an ICMP request as soon as it receives a response. The results of this scenario are provided in Table 3.

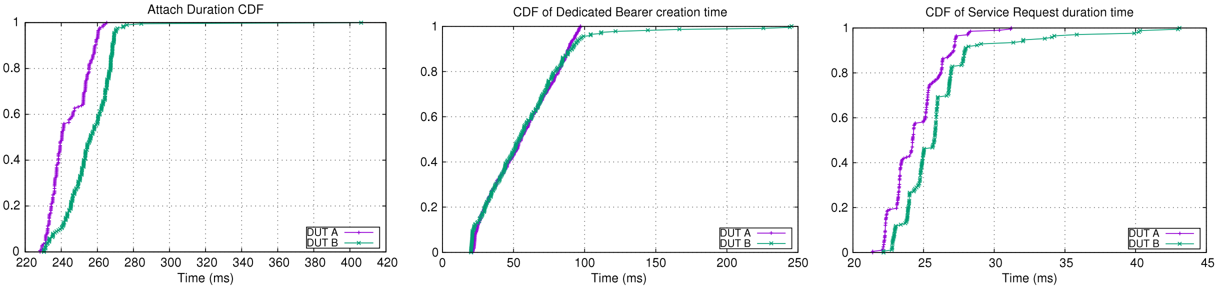
Additionally a comparison between the ideal and non-ideal channels is depicted in Fig. 3. The figure provides the cumulative distribution function for the end-to-end delays. The behaviour in the ideal channel is better for both DUTs; there are no radio access retransmissions so all the packets arrive in under 100 ms. For the non-ideal channel case a vehicular scenario with an SNR of 6 has been considered. When introducing the fading and noise the results are slightly worse, with more samples that exceed the 100 ms RTT.



(a) Ideal channel.

(b) Non ideal channel.

Fig. 3. Data plane latency measurements.



(a) CDF attach.

(b) CDF DRB establishment.

(c) CDF service request.

Fig. 4. Control plane latency measurements.

4.5. *Latency of the control plane*

The control plane latency can also affect the overall data plane delay of the system. For instance the attach procedure will determine how fast the devices will be ready to send data when powered off, the service request will indicate how fast the modem is ready to send data after exiting from an idle state and the dedicated bearer establishment time will indicate the time consumed to start to prioritize data, as described in Section 5.2.

To characterize the behaviour of the UEs the EPC emulators have been collocated with the base stations and certain scripts have been implemented. To produce attach samples, the EPC has been commanded to trigger a detach request with a reattach required indicator. The service request has been achieved by commanding the UE to release the context and the dedicated bearer establishment by forcing an establishment and a release continuously.

Fig. 4 depicts the CDF for the three procedures analysed. The dedicated bearer establishment consumes approximately the same time in both devices while the attach and service request procedures are completed in less time by the DUT A. Table 4 provides a summary of the median and MAD time for each of the DUTs. It is important to note that these results are much lower than in real networks, the main reasons for this are the absence of other devices accessing the cell and the local deployment of the core network. The local deployment will introduce an additional delay on all the procedures due to the transport of the information from the base station to the EPC. The lack of other devices will worsen the attach time, which will be higher due to larger random access delays, and the dedicated bearer procedures, which could fail due to lack of resources in the base station. However, for the sake of comparison the results enough as we remove part of the uncertainty that could be introduced by external factors.

5. The network optimization perspective

Besides the UE themselves the network can also be optimized to support mHealth applications. Some of the requirements typically required by these types of applications are high priority, low latency communication (e.g.: for real-time communications), group communications (sensor networks) or traffic prioritization (critical applications). In this section a

Table 4
Summary of the control plane procedures.

Procedure	DUT A Median (MAD) time	DUT B Median (MAD) time
Attach	240.35 (9.29) ms	256.34 (10.39) ms
Dedicated Bearer Est.	54.06 (20.56) ms	55.49 (23.6) ms
Service Request	24.225 (1.5) ms	25.78 (2.076) ms

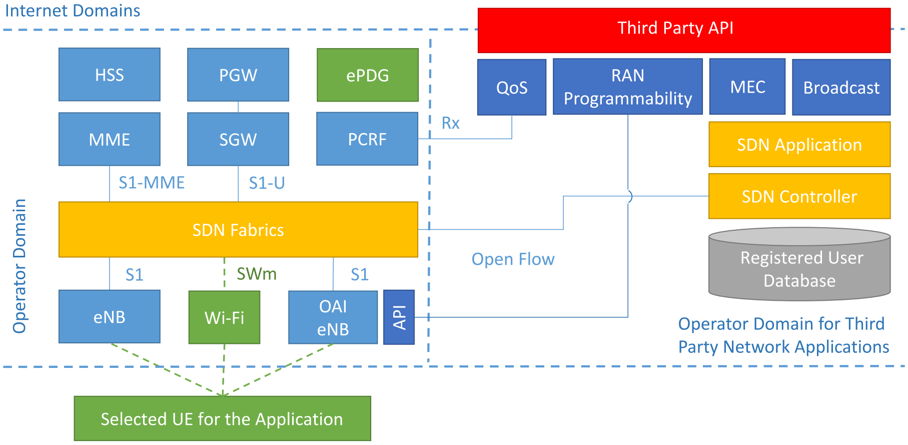


Fig. 5. The Q4Health project's overall architecture.

full architecture supporting these features is presented and the different functionalities and components are described in detail.

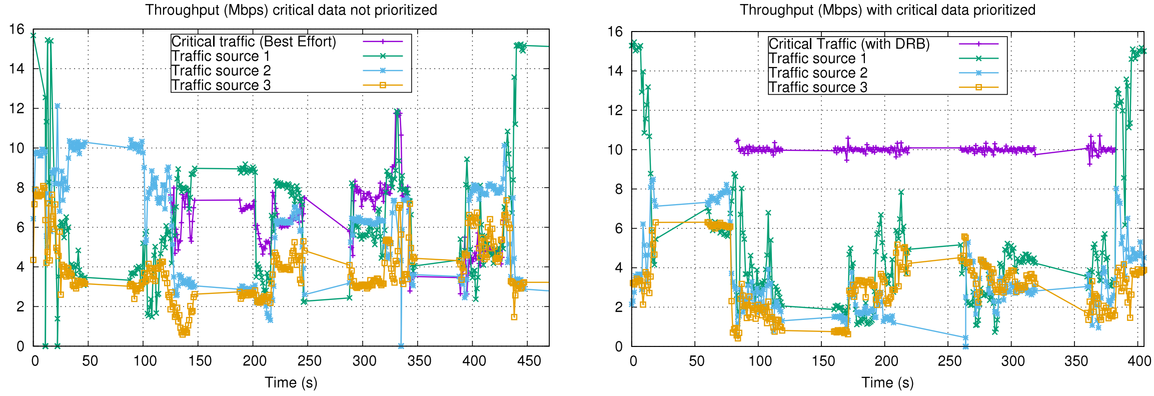
5.1. Network architecture for mHealth applications

Fig. 5 depicts an overview of the proposed architecture. There are three different areas, each related to different domains (operator, operator for third party and Internet). The mobile network equipment (mainly the different radio access technologies and the core network) is part of the operator domain, and an operator domain for third parties is also considered. This domain contains the components which affect the network that could be deployed by third parties. Finally the different Internet domains are also considered. A specific interface (the third party API) to expose network functions to mHealth applications has also been considered.

In the operator domain the LTE architecture is the standard one, with new components added to support seamless handover to non 3GPP technologies. The main functionality is provided by the evolved Node B (eNB), which is the radio base station of LTE and evolved Packet Core (EPC), which is the core network in LTE. In the EPC the basic functionality is provided by the Mobility Management Entity (MME), the Serving Gateway (SGW), the Packet Data Network (PDN) Gateway (P-GW), the Home Subscriber Server (HSS) and the Policy Charging and Subscriber Function (PCRF). The MME is in charge of the control plane, supporting procedures such as registration (attach), dedicated bearer establishment or handover, and is connected to the radio stations using the S1-MME interface. The P-GW and S-GW can be deployed together and provides functionality for routing and forwarding traffic tunnels to external data networks. The SGW is connected to the eNB using the S1-U interface. HSS provides information about the users (subscription details, security, etc.), while the PCRF is in charge of guaranteeing QoS rules in the system. This component also includes an interface for third parties to produce QoS demands, which can be used in real-time.

The operator domain also includes a functionality to support non-3GPP technologies, such as Wi-Fi, which is provided by the Access Network Discovery and Selection (ANDSF), responsible for assisting the UE in the discovery of valid networks, and the Evolved Packet Data Gateway (ePDG), which is designed to secure the connection from untrusted non 3GPP networks connected to the EPC. These components can improve the interaction with heterogeneous technology, very frequent in mHealth applications, and improve the behaviour in indoor deployments, which can have low mobile signal level but have Wi-Fi available.

For the connectivity of the network components SDN fabric elements have been considered, which are compatible SDN L2/L3 switching elements. The use of SDN switches enables forwarding and/or redirecting packets to an SDN application, which is used to enable low latency services and group communications, as described in Section 5.3. The rules to match the packets are configured by the SDN application, which is running on top of an SDN controller.



(a) Traffic comparison without QoS.

(b) Traffic comparison with QoS.

Fig. 6. Traffic comparison between multiple UEs with and without prioritization.

Table 5
Summary of the results for the different UEs.

	Throughput Best Effort (Mbps)	Throughput Prioritization (Mpbs)
UE Background Traffic 1	6.210	5.402
UE Background Traffic 2	6.077	3.176
UE Background Traffic 3	3.89	3.078
UE Critical Traffic	6.351	10

In our approach the functionality to support real-time requirements on the network resources is offered by the third party API. This API is based in on the VELOX system, which is a virtual path slice engine that is used to manage bandwidth requests from third party applications. The system exposes functionality to enforce QoS across domains, to request bandwidth and latency characteristics from the network and to pass information to the radio access schedulers about the type of traffic being managed.

In Fig. 5 the API functionality is depicted using different blocks. The QoS block provides access to the Rx functionality of the PCRF, which can be used to trigger dedicated bearers with better QoS characteristics, and QoS enforcement in the operator SDN path. The RAN programmability block, described in Section 5.5, offers access to the scheduler in the LTE MAC (Medium Access Control) layer of the eNB and the MEC and Broadcast modules are used to provide low latency services and group communications respectively. Due to the provision of MAC layer enhancements the system is able to build over-the-top traffic prioritization and QoS mechanisms providing better adaptation to applications than regular schedulers.

5.2. Third party QoS demands

The QoS API module provides user a way of accessing the functionality in the Rx interface to setup dynamically certain QoS for an specified type of traffic. Once the user requests a certain quality the API first creates a dedicated bearer to protect the control plane of the application (for instance for video or audio the Session Initiation Protocol (SIP) messages could be protected) and then a bearer for the actual traffic is requested from the network. The LTE standard defines different Quality Class Indicators (QCI), which are defined in [34].

The approach has been validated by connecting four terminals to a base station and executing different tests with and without triggering a level of QoS. Three of the UEs are employed to generate background traffic, in order to fill all the available bandwidth in the uplink they generate a constant bitrate of 30 Mbps. The other UE is considered the one to be generating critical traffic at a rate of 10 Mbps.

Fig. 6 depicts the bandwidth estimation for the uplink of the four terminals when (a) the critical traffic is sent under best effort conditions, and (b) when the critical traffic is prioritized using the VELOX interface. When all the flows are sent in best effort mode the throughput is divided more or less equally (three of the UEs around 6 Mbps and one of them in 3Mpbs, which is probably due to problems in the antenna leading to bad channel conditions), the mean aggregated throughput for the background traffic is 16.18 Mbps. Once the prioritization for one the critical traffic is applied the mean background traffic is reduced to 11.657 Mpbs as the base station start to grant more frequently the UE generating the critical data.

A summary of the results for each of the UEs is provided in Table 5. Fig. 7 depicts the comparison between the critical traffic being transported with and without triggering a QoS request to the third party API. If the bearer is installed correctly the traffic is completely transported.

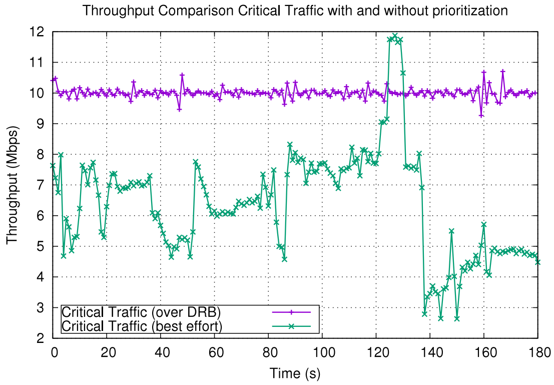


Fig. 7. Comparison between traffic transported over default and dedicated bearer.

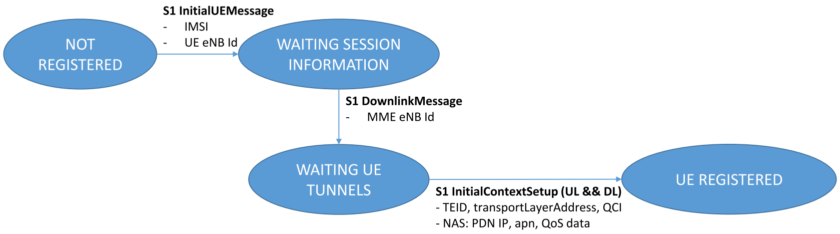

NOT REGISTERED
S1 InitialUEMessage
- IMSI
- UE eNB Id
WAITING SESSION INFORMATION
S1 DownlinkMessage
- MME eNB Id
WAITING UE TUNNELS
S1 InitialContextSetup (UL && DL)
- TEID, transportLayerAddress, QCI
- NAS: PDN IP, apn, QoS data
UE REGISTERED

Fig. 8. Database building state machine.

5.3. Low latency services

In [6] and [7] a first analysis of the effect of a system exploiting the LTE data plane to reduce the end-to-end latency is given. In these two papers we explored the provision of reduced latency fog services by introducing an intermediate component that is in charge of building a tunnel database by analysing the data plane packets. This information can be used to redirect the traffic from and to the fog, avoiding the delay introduced by the EPC. The preliminary results, described in the papers, showed a reduction of up to 78% in the RTT of the Fog Traffic. This approach works well in scenarios with low mobility but when a high rate of handovers is considered, the improvements in performance are minor. Furthermore good results require traffic in both the uplink and the downlink, without this the system is unable to build the required database.

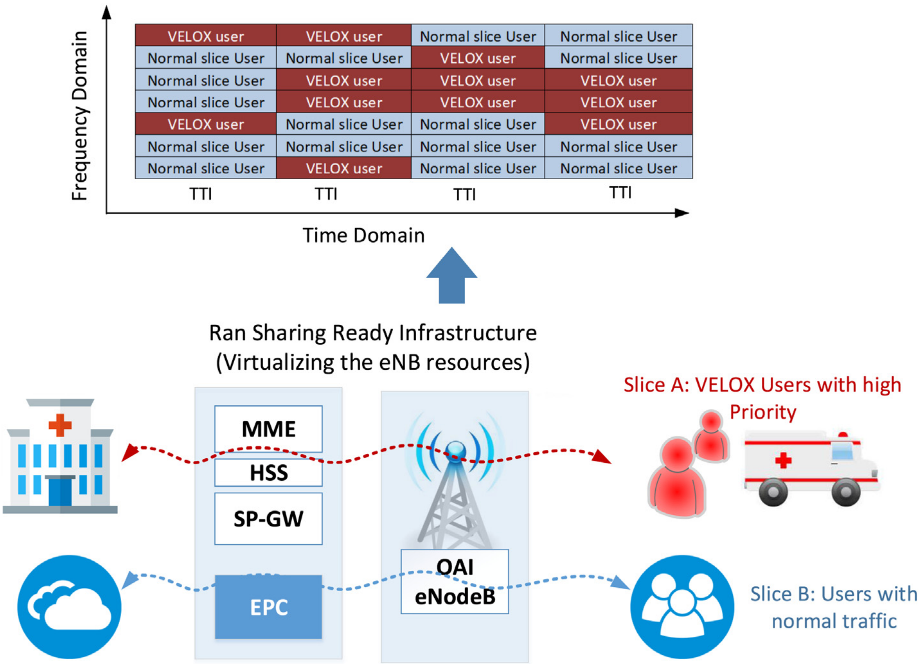
To mitigate these effects another approach is being explored which consists in analysing the control plane. A sniffer capable of analysing the traffic of the S1-MME interface has been built. This sniffer is able to build a complete database before any data has been exchanged in the data plane thus removing the requirement of having data in both directions. Again this approach can be implemented with an SDN application, which will receive all the control plane data to build the tunnel database and, based on this information, will push rules to the SDN fabrics in order to decide whether the traffic should be routed towards the EPC or not.

Fig. 8 depicts the state machine that is used by the implemented software to generate the database. The Ids of the UEs are stored in order to keep track of the connection until the procedure is completed. The main elements to build the database the IMSI, which is to determine if the user is authorized to access the fog, and the Tunnel Identifier Endpoint (TEID) and the transportLayerAddress for uplink and downlink, which is the actual information of the tunnel. The NAS information can only be extracted if security is disabled so normally this is not available. The most important part of the information is the UE PDN IP, which can also be inferred from the analysis of the data plane.

5.4. Group communications

The aforementioned approaches can be employed to provide an efficient mechanism to support multicast communications between a group of users. The SDN switches, which sit between the radio access and the core network, can be employed to duplicate packets towards the different peers of the group. For peers connected to base stations belonging to the same SDN application the latency will be reduced and for peers outside, the packets will be forwarded more efficiently as there is no need to send to a distribution peer.

In an LTE standard network there are components devoted to group communication (namely the eMBMS architecture, described in [35]), which introduces a new physical channel for broadcasting. These standard elements are designed for



VELOX user	VELOX user	Normal slice User	Normal slice User
Normal slice User	Normal slice User	VELOX user	Normal slice User
Normal slice User	VELOX user	VELOX user	VELOX user
Normal slice User	VELOX user	VELOX user	VELOX user
VELOX user	Normal slice User	Normal slice User	VELOX user
Normal slice User	Normal slice User	Normal slice User	Normal slice User
Normal slice User	VELOX user	Normal slice User	Normal slice User
TTI	TTI	TTI	TTI

Frequency Domain
Time Domain
Ran Sharing Ready Infrastructure
(Virtualizing the eNB resources)
MME
HSS
SP-GW
EPC
OAI
eNodeB
Slice A: VELOX Users with high Priority
Slice B: Users with normal traffic

Fig. 9. Different network slices operating over shared eNodeB.

the distribution of content among large groups of users, while IoT for mobile health group communications are normally demanded for smaller groups. The MBMS services transmit content over one are and a base station can belong to eight of these areas. A possibility to exploit the best of both approaches could be the use of one of this areas to be used dynamically by mHealth services requiring group communications.

5.5. *RAN programmability*

5.5.1. *Existing scheduling approaches in LTE*

In LTE networks downlink transmissions are grouped in (radio) frames 10 ms long, one radio frame is formed of 10 subframes of 1 ms duration and there are ten subframes in the uplink and ten frames in the downlink. Each subframe is divided into two slots of 0.5 ms duration. Each slot counts 6 or 7 OFDM symbols for normal or extended cyclic prefix used. The Physical Resource Block (PRB) is the smallest element assigned by the base station scheduler. Transmission Time Interval (TTI) is the duration of a transmission on the radio link. The TTI is related to the size of the data blocks passed from the higher network layers to the radio link layer. A scheduler can determine to which user the shared resources (time and frequencies) for each TTI (1 ms) should be allocated.

Although many approaches consider optimal PRB allocation in order to maximize throughput by means of service differentiation between the user, like for example in [17], the approach is also proportionally fair, using a joint optimization framework [36,37]. Note however that actually there is no application-driven policy enforcement or service differentiation between groups of users.

5.5.2. *Policy based scheduling and priority in the uplink*

The objective of the proposed RAN programmability framework is to implement policies that are able to exploit SDN principles and achieve over-the-top service differentiation. In the solution devised, the system is able to schedule Resource Blocks (RBs) effectively between different groups of users with respect to specific QoS objectives and isolation guarantees between the different users and types of traffic. Fig. 9 depicts a visual representation, where the scheduling principle will decide which PRBs will be allocated to specific UEs according to the signed quality service class.

The resource allocation by means of Resource Block scheduling is based on feedback received from the UE, existing allocations and the SLAs signed. This feedback is based on the Channel Quality Indicator (CQI) that contains information sent from an UE to the eNodeB to indicate a suitable downlink transmission data rate, i.e. a Modulation and Coding Scheme (MCS) value (CQI is a 4-bit integer and is based on the observed signal-to-interference-plus-noise ratio (SINR) at the UE). In principle, prior to the Resource Block allocation, the UE reports to the eNodeB which radio bearers (or Logical Channel) need uplink resources and how many resources they need. Each logical channel is associated with a service belonging to a specific QoS class. Depending on the service requests, there is a buffer queue in the user's respective logical channel. Each

Table 6
Buffer queue is identified by a set of parameters.

Packet-level parameters
• Average packet size (APS)
• Packets inter-arrival time (PIT)
• Packet arrival time (PAT)
• Packet maximum-allowable delay (PMD)
• Packet remaining time (PRT)
Logical channel-level parameters
• Number of protocol data units (NPDU)
• Head-of-line delay (HOD)
• Buffer size of a logical channel (BS)
• Guaranteed bit-rate (GBR)
• Level of traffic (TL)
User-level parameters
• Total buffer size (TBS) • Number of logical channels (NLC)
• Channel quality indicator (CQI)
• Subscription type (ST)
System-level parameters
• Number of active users (NU)
• Maximum allowed scheduled users (MAU)
• Frame configuration type (CT)
• Transmission time interval (TTI)
• Minimum resource allocation unit (MRU)

buffer queue is identified by a set of parameters that are used to design an optimal scheduling algorithm. The final scheduling decision is based on the QoS characteristic of the corresponding radio bearers and the reported buffer status. In Table 6 the most important parameters that can be used to affect the scheduling are presented, spanning from packet level and system level to user level and logical channel parameters.

5.5.3. Enabling programmability in OAI eNodeB

In order to determine the necessary scheduling approach that will be used to facilitate the change in prioritization, we exploit an SDN-based approach where a logically centralized Controller will be used to instruct a local agent at the eNodeB for the scheduler in effect. All the proof of concept demonstrations are performed using the open-source OAI LTE platform [38]. The following approach is considered:

- Use SDN principles to address the challenges in the RAN.
- Separation of the control and the data plane.
- Centralized view of the RAN for improved decisions.
- Re-programmability of the data plane on the fly by a centralized controller.

Essentially, the controller will be connected to a number of eNodeBs, and will be able to directly control and modify their states. All the significant control logic will be removed from layers L2/L3 and will be integrated to a higher level controller. The extracted control logic will be the scheduler of the MAC layer or the mobility management function of the RRC layer. The scheduling policies we design will be able to consider the necessary traffic characteristics to meet the Q4HEALTH QoS application requirements by means of jitter, latency, data rate, and loss rate as well as fairness between end users.

The standard separation of the control and the data plane is used, however time critical zones are considered for the re-programmability of the data plane on the fly by the centralized controller. In our solution we also take into account the need for partial or full control delegation by the controller over the agent, based on run-time conditions. Such delegation could be, for example, decision making about mobility management (RRC layer). Another important feature is the ability to reprogram the functionality of agents; for example, we are able to write and load a new scheduler at the eNodeB based on the current network load.

Note that by following this SDN design a standard protocol is required for the Agent–Controller communication. For the SDN eNodeB programmability approach we will exploit and extend the work delivered and presented in [39]. The implementation details, performance analysis and a novel southbound protocol called FlexProtocol for the Controller–Agent communication can be found therein.

5.5.4. Framework design

For the scheduler framework design itself a three layer scheduler implementation (MAC-layer scheduling) proposed in [40] is used to provide a more dynamic approach to satisfy the operator-driven, technology driven and service-driven requirements through re-configurable APIs. The framework is used for the downlink but it will be extended to also support the uplink. This design allows the provisioning of the required QoS to every client, while differentiating services and prioritizing traffic between VELOX users and other stakeholders; essentially scheduling the RB in such a way that third party applications receive the agreed service quality level. The scheduler programmability will be supported by the real-time agent functions.

Please cite this article in press as: C. Garcia-Perez, et al., Improving the efficiency and reliability of wearable based mobile eHealth applications, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.06.021.

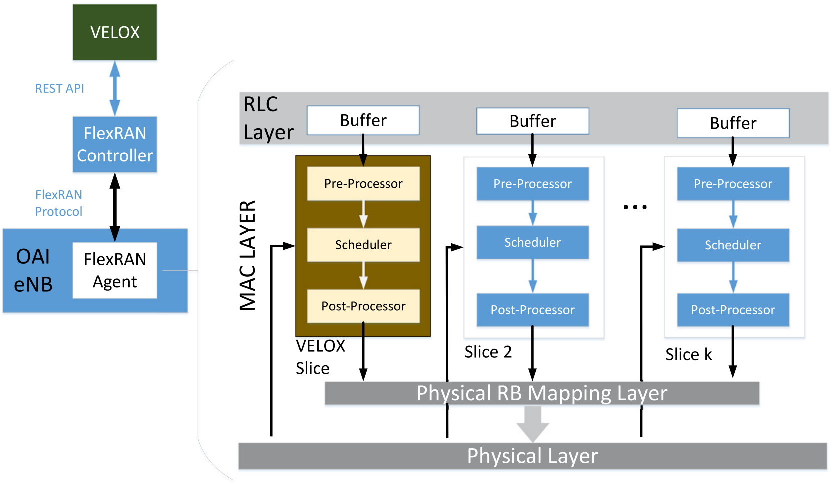


Fig. 10. eNB virtualization and network slicing in the OAI eNB MAC layer.

The framework is used to support both the downlink and the uplink operations. This design allows the provisioning of the required QoS to every client, while differentiating services and prioritizing traffic between VELOX users and other stakeholders; essentially scheduling the RB in such a way that third party applications receive the agreed service quality level. The scheduler programmability is supported by the real-time FLEXRAN [39] agent functions .

The framework structure in terms of its interfaces and modules of MAC-layer is shown in Fig. 10. The framework is able to provide different Preprocessor–Scheduler and Post processor on per Network Slice basis. This way every different slice (e.g. VELOX) is able to exploit its own scheduling principle and satisfy the design goals, without affecting the operation of the other slices. Every slice assumes operation with no other slice operation and the concept of Virtual Resource Blocks is used. This way in the MAC layer, every user is scheduled by his slice owner assuming no other slice operation. A physical RB Mapping layer is responsible for the mapping of the virtual resource blocks to actual physical resource blocks. This layer is also responsible to guarantee isolation between slices and specific shares on the frequency and time domain.

For every slice the main software components of the architecture are the following:

- MAC-Buffer: This interface shares the packet information of users and logical channels with the MAC-layer. Each logical channel represents a service with specific values of KPI parameters.
- MAC-PHY: Most of the modern scheduling algorithms are channel-aware, therefore it is crucial for the MAC-layer to have a knowledge of the channel quality information of all the active users in the system. Once the channel estimation has been done in the user terminals, the result is sent to the base station via feedback channel. The physical layer receives this information and forwards it through the MAC-PHY interface to MAC. Channel quality information is used while calculating the expected throughput for a user and in turn for the entire system.
- Pre-processor: This module represents a novel extension to the traditional scheduling framework. The main function of the pre-processor is to convert the two-dimensional buffer of users logical channels into a single dimension vector.
- Scheduler: Once the conversion to a single dimension has been done in the pre-processor, the first task of the scheduler module is to deal with sorting of blocks based on scheduling requirements. For example, in the round robin case, the blocks remain in the order of their index. After block sorting, the actual allocation of resources to the sorted blocks is carried out.
- Post-processor: The post-processor implementation depends on the wireless standard. The mapping of users to resource elements in the frequency domain is applied, based on the system specifications. For example, in 3GPP LTE, the selection of Resource Block Groups (RBGs) for a particular service of a user is done by the post-processor. Note that the number of RBGs depends on the available bandwidth.

With this approach the satisfied GBR percentage per user, as well as the system throughput improves for all the traditional scheduling algorithms.

5.5.5. Framework evaluation

The evaluation of our proposal is done using OAI with a third party EPC and two user equipment. For simplicity, we consider the case that two users belong to different slices. One slice is dedicated to the VELOX users where we want to dynamically adjust the resource block share and affect the end-user performance. We use the LTE FDD SISO mode with 5MHz bandwidth and measure our interested metrics with uplink UDP traffic. Firstly, we examine the per-slice policy on the frequency domain resource allocation (i.e., Physical Resource Block (PRB) in LTE) in two different ways.

Please cite this article in press as: C. Garcia-Perez, et al., Improving the efficiency and reliability of wearable based mobile eHealth applications, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.06.021.

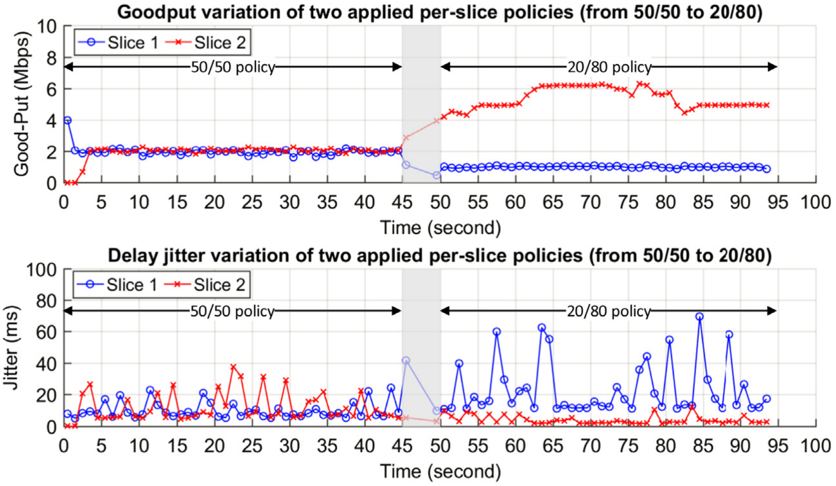


Fig. 11. Apply time-varying per-slice priority policy.

1. Fair slice policy: It refers to the case that each slice has the same priority and share the same portion of all available resource. In following, it is denoted as 50/50 policy.
2. Prioritized slice policy: In contrast, different priorities are put into practice on different slices. For instance, the first slices are more favoured and can allocate 80% of all available resources whereas another slices only allocate the rest 20% of resources. It is denoted as 20/80 policy in following result. The higher percentage goes to the case of the eHealth (VELOX) users.

In Fig. 11, we show the result of applying these two policies in a time-varying manner. These results are measured at the application-layer in terms of good-put and delay jitter. In the beginning 45 s, we apply the 50/50 policy but change into 20/80 policy within the duration from 45 s to 50 s. We can observe that the goodput will be significantly changed after applying the prioritized slice policy. Further, the delay jitter is also impacted due to a low-prioritized slice can only use fewer resource blocks and increase the application-layer delay variance.

In following, we further extend the per-slice policy to incorporate the per use-case policy. Such per use-case policy can be applied with different 5G use-cases [2], e.g., Extreme Mobile BroadBand (xMBB), ultra-Reliable Low Latency Communication (URLLC) in order to satisfy the different performance requirements. In the context of the Q4Health solution we are highly interested in the case of ultra-Reliable Low Latency Communication (URLLC) case. We introduce and evaluate over the RAN Sharing framework two policies are introduced as follows:

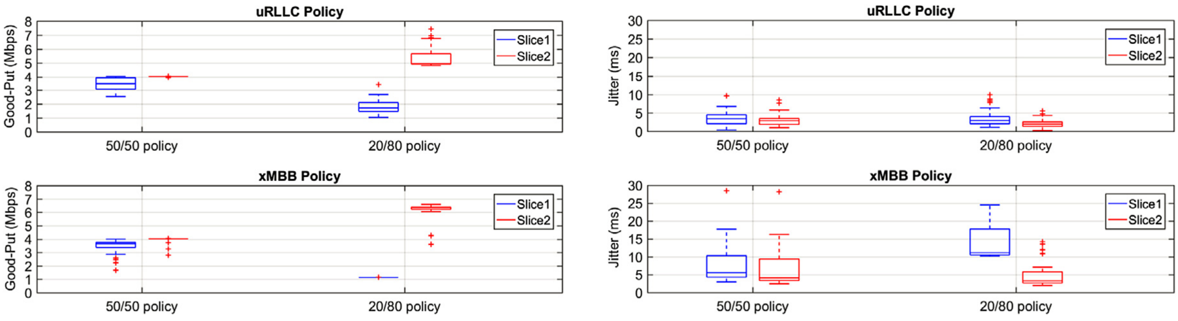
1. xMBB policy: A more extreme modulation and coding scheme (MCS) index is allocated in terms of the extra positive offsets based on the reported wideband channel quality indicator (CQI) in order to further enhance the data throughput but may increase the number of re-transmissions. In following, the offset level of MCS index is set to 2.
2. uRLLC policy: In this case, a more conservative MCS index is allocated based on the reported CQI value. Further, the UE-selected sub-band CQI report can be applied to increase the reliability and reduce the delay variation due to fewer re-transmissions.

In Fig. 12, the results are presented in terms of goodput and delay jitter in terms of joint slice and use-case policy. We can observe that the xMBB policy is superior to the uRLLC one in terms of goodput Fig. 12(a) in most of the cases except the slice 1 of 20/80 policy. This suggests that the extreme MCS allocation may not suitable for the de-prioritized slices. However, in terms of delay jitter of Fig. 12(b), uRLLC shows a lower variation compared with the one of xMBB policy of all cases. Propitiate policies can be applied for different use-cases in order to distinguish their advantages.

6. Discussion

6.1. UE selection results

The two DUTs analysed offer different characteristics. For instance, DUT B demonstrates a better performance in the cell edges and also higher throughput, this device will perform well for bandwidth demanding applications with high mobility. On the other hand DUT A offers better delays and lower costs, which will be positive for a delay sensitive application. For this particular scenario power consumption has been similar but in general it will differ proportionately, inversely to the throughput characteristics.



(a) Goodput variation.

(b) Delay jitter variation.

Fig. 12. Comparison of different use cases.

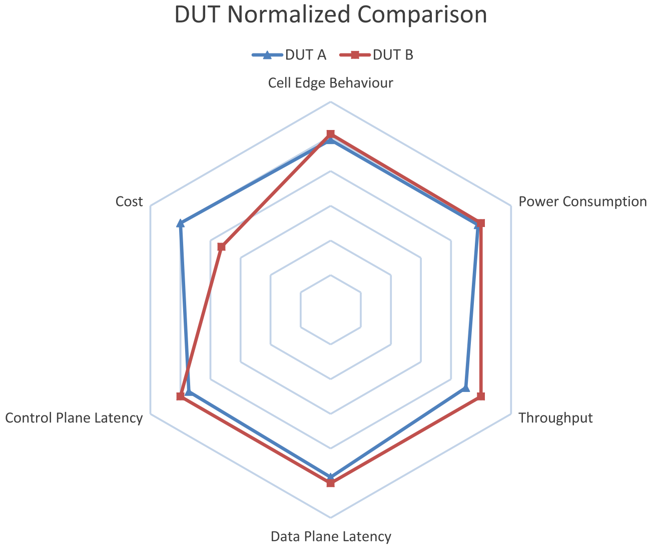


Fig. 13. Normalized Comparison of the results for DUT A and DUT B.

Fig. 13 provides a comparison of all the aspects considered, the values have been normalized to the maximum to ease the comparison. Cell edge behaviour has been calculated as the mean between the normalized values of the receiver sensitivity and the maximum power. For the control plane latencies a normalized mean has also been provided but a specific metric for an application could be defined (e.g.: better service request). The cost has also been introduced and others physical parameters could also have been included (e.g.: weight, dimensions, etc.) if relevant for the application. For our particular use case DUT A was selected as the latency and power consumption were very important for the platform.

The set of measurements defined by our approach target different requirements of mHealth applications. Normally an application will only require a subset of these measurements to work better so obtaining the values for different devices can help to assess which of them perform better for the desired application. Additionally using the devices in their final platform also provides a more realistic view of the actual behaviour as it will include the effects of the driver, the operating system, the data patterns, etc. On the other hand this approach could not be applied to all the applications as for many of them require accessing the antenna ports of the modem, which are always not accessible.

6.2. Network optimization results

The proposed API can accelerate the deployment of mHealth applications by simplifying the exposure of network functions. We have not taken into account the security aspects which are critical both for the domain of the application and for the operators implementing this solution. The business model and commercialization should also be explored in future works.

The use of dedicated bearers considerably improves the behaviour of the traffic, which stop suffering packet loss (in our test moving from a 37% to a 0%) due to saturation and maintain better figures of throughput. The effect of the number of users on the cell has to be evaluated in order to determine the limits for the number of users that can cause establishment failures.

Our initial approach to reduce latency has shown promising results, and the modifications suggested in this paper will improve the behaviour in scenarios without downlink traffic to cloud services. The very same approach can also be exploited to support group communications to reduce the overhead and the latency for part of the members of the group. Both approaches are still under development and require a quantitative analysis.

Finally the RAN programmability has shown promising results for the downlink. Our future plans include all the necessary extensions to support the relevant traffic prioritization for the uplink. Efficient resource allocation policies and MAC scheduling principles will be exploited in order to provide the necessary QoS characteristics for the SLAs that external mHealth applications sign for.

7. Conclusion

In this paper we have presented improvements in meHealth applications thanks to network capabilities that are already deployed albeit rarely used. These include the real-time configuration and use of traffic prioritization, and the emergent paradigms of SDN and NFV, with the reliability needed to be used in the field of security and emergency services.

We have presented an approach, applied to a wearable video application for emergency services, to select the most appropriate modem and validate the functionality by the comparison of two different devices for the use case. The results shown that the best device to use will indeed depend on the requirements of the application. From a network point of view we have designed an API to improve and accelerate the deployment of critical mHealth applications. The API is able to trigger QoS requests, which in our tests has increased the throughput in a 58%; setup low latency fog services, which can achieve a latency reduction of the 78%, setup group communications, and support fine grain scheduling at the base station traffic scheduler.

Although this project focuses on mHealth applications, with clear requirements of stability and quality assurance, the improved efficiency and latency reduction we hope to achieve in the system as a whole will also be very useful in a world full of IoT devices. This type of network user, which most analysts claim will be omnipresent in the near future, needs to share the limited available resources with thousands of others, who will welcome every step towards increasing the availability and stability of the operator network.

The next step, and arguably the most challenging one, will be the integration of very different technologies in all the layers of the mobile network stack, all of them synchronized and working seamlessly. After that phase is complete a tuning effort will be required to extract the best performance from each component and study the limits of the deployment to extrapolate the results to be used in live and commercial networks where the stability of the system is a must.

Acknowledgement

The work described in this paper has received funding from the European Horizon 2020 Programme under grant agreement 688624 (Q4Health project).

References

[1] W.H. Organization, mHealth: New horizons for health through mobile technologies: second global survey on eHealth, World Health Organization, 2011. URL http://www.who.int/goe/publications/goe_mhealth_web.pdf.

[2] F.A. Kraemer, A.E. Braten, N. Tamkittikhun, D. Palma, Fog computing in healthcare; a review and discussion, IEEE Access PP (99) (2017) 1. http://dx.doi.org/10.1109/ACCESS.2017.2704100.

[3] E. Kartsakli, A.S. Lalos, A. Antonopoulos, S. Tennina, M.D. Renzo, L. Alonso, End-to-end communication challenges in M2M systems for mHealth applications, in: 2014 IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD, 2014, pp. 355–359. http://dx.doi.org/10.1109/CAMAD.2014.7033265.

[4] S. Forge, R. Horvitz, C. Blackman, Is commercial cellular suitable for mission critical broadband? Study on use of commercial mobile networks and equipment for "mission-critical" high-speed broadband communications in specific sectors Tech. rep., European Union,2014 .URL http://dx.doi.org/10.2759/54788.

[5] C.A. Garcia-Perez, A. Rios, P. Merino, K. Katsalis, N. Nikaein, R. Figueiredo, D. Morris, T. O'Callaghan, Q4HEALTH: Quality of Service and prioritisation for emergency services in the LTE RAN stack, in: 2016 European Conference on Networks and Communications (EuCNC), 2016, pp. 64–68. http://dx.doi.org/10.1109/EuCNC.2016.7561006.

[6] C.A. Garcia-Perez, P. Merino, Enabling low latency services in standard LTE networks, in: Foundations and Applications of Self-* Systems (FAS*), 2016 IEEE 1st International Workshops, 2016, pp. 248–255. http://dx.doi.org/10.1109/FAS-W.2016.59.

[7] C.A. Garcia-Perez, P. Merino, Experimental evaluation of fog computing techniques to reduce latency in LTE networks, Trans. Emerging Telecommun. Technol. (2017) in press, http://dx.doi.org/10.2759/54788 http://dx.doi.org/10.1002/ett.3201.

[8] A. Diaz, C.A. Garcia-Perez, A. Martin, P. Merino, A. Rios, PerformNetworks: a testbed for exhaustive interoperability and performance analysis for mobile networks, in: Building the Future Internet Through FIRE, River Publishers, 2017, pp. 1–250.

[9] A. Díaz-Zayas, C.A. García-Pérez, Á.M. Recio-Pérez, P. Merino, PerformLTE: a testbed for LTE testing in the future internet, in: Wired/Wireless Internet Communications: 13th International Conference, WWIC 2015, Malaga, Spain, May 25–27, 2015, Revised Selected Papers, Springer International Publishing, Cham, 2015, pp. 46–59. http://dx.doi.org/10.1007/978-3-319-22572-2_4.

[10] F. Kaltenberger, R. Knopp, N. Nikaein, D. Nussbaum, L. Gauthier, C. Bonnet, OpenAirInterface: Open-source software radio solutions for 5G, in: EUCNC 2015, European Conference on Networks and Communications, 29 June–02 July 2015, Paris, France, 2015. URL http://www.eurecom.fr/publication/4634.

[11] L.K. Moore, The First Responder Network (FirstNet) and Next-Generation Communications for Public Safety: Issues for Congress, Congressional Research Service, 2016 URL https://www.fas.org/sgp/crs/homesec/R42543.pdf.

[12] A. Prasad, A. Maeder, K. Samdanis, A. Kunz, G. Velev, Enabling group communication for public safety in LTE-Advanced networks, J. Netw. Comput. Appl. 62 (2016) 41–52. http://dx.doi.org/10.1016/j.jnca.2015.10.014.

[13] T.-T. Nguyen, C. Bonnet, N.-D. Nguyen, LTE Broadcast for Public Safety, in: D. Cmara, N. Nikaein (Eds.), Wireless Public Safety Networks 2, Elsevier, 2016, pp. 263–293. http://dx.doi.org/10.1016/B978-1-78548-052-2.50009-8. URL http://www.sciencedirect.com/science/article/pii/B9781785480522500098.

[14] R. Favraud, A. Apostolaras, N. Nikaein, T. Korakis, Public Safety Networks: Enabling Mobility for Critical Communications, in: D. Cmara, N. Nikaein (Eds.), Wireless Public Safety Networks 2, Elsevier, 2016, pp. 95–126. http://dx.doi.org/10.1016/B978-1-78548-052-2.50004-9.

[15] N. Maskey, S. Horsmanheimo, L. Tuomimki, Latency analysis of LTE network for M2M applications, in: Telecommunications (ConTEL), 2015 13th International Conference on, 2015, pp. 1–7. http://dx.doi.org/10.1109/ConTEL.2015.7231227.

[16] S. Fouziya Sulthana, R. Nakkeeran, Study of downlink scheduling algorithms in LTE networks, Journal of Networks 9 (12) (2014) 3381–3391. http://dx.doi.org/10.4304/jnw.9.12.3381-3391.

[17] N. Abu-Ali, A.E.M. Taha, M. Salah, H. Hassanein, Uplink Scheduling in LTE and LTE-Advanced: Tutorial, survey and evaluation framework, IEEE Commun. Surv. Tutor. 16 (3) (2014) 1239–1265. http://dx.doi.org/10.1109/SURV.2013.1127.00161.

[18] A. Bhamri, N. Nikaein, F. Kaltenberger, J. Hmlinen, R. Knopp, Pre-processor for MAC-layer scheduler to efficiently manage buffer in modern wireless networks, in: 2014 IEEE Wireless Communications and Networking Conference, WCNC, 2014, pp. 1544–1549. http://dx.doi.org/10.1109/WCNC.2014.6952439.

[19] T. Erpek, A. Abdelhadi, T.C. Clancy, Application-aware resource block and power allocation for LTE, in: 2016 Annual IEEE Systems Conference (SysCon), 2016, pp. 1–5. http://dx.doi.org/10.1109/SYSCON.2016.7490591.

[20] M.S. Mushtaq, S. Fowler, A. Mellouk, B. Augustin, QoE/QoS-aware LTE downlink scheduler for VoIP with power saving, J. Netw. Comput. Appl. 51 (2015) 29–46. http://dx.doi.org/10.1016/j.jnca.2014.02.001.

[21] R. Kwan, C. Leung, J. Zhang, Downlink Resource Scheduling in an LTE System, INTECH Open Access Publisher, 2010. http://dx.doi.org/10.5772/7687.

[22] V.-G. Nguyen, T.-X. Do, Y. Kim, SDN and Virtualization-Based LTE Mobile Network Architectures: A Comprehensive Survey, Wirel. Pers. Commun. 86 (3) (2016) 1401–1438. http://dx.doi.org/10.1007/s11277-015-2997-7. URL http://dx.doi.org/10.1007/s11277-015-2997-7.

[23] J. Heinonen, T. Partti, M. Kallio, K. Lappalainen, H. Flinck, J. Hillo, Dynamic tunnel switching for SDN-based cellular core networks, in: Proceedings of the 4th Workshop on All Things Cellular: Operations, Applications, &#38; Challenges, in: AllThingsCellular'14, ACM, New York, NY, USA, 2014, pp. 27–32. http://dx.doi.org/10.1145/2627585.2627587. URL http://doi.acm.org/10.1145/2627585.2627587.

[24] L. Hu, M. Qiu, J. Song, M.S. Hossain, A. Ghoneim, Software defined healthcare networks, IEEE Wirel. Commun. 22 (6) (2015) 67–75. http://dx.doi.org/10.1109/MWC.2015.7368826.

[25] J. Santos, J.J. Rodrigues, B.M. Silva, J. Casal, K. Saleem, V. Denisov, An IoT-based mobile gateway for intelligent personal assistants on mobile health environments, J. Netw. Comput. Appl. 71 (2016) 194–204. http://dx.doi.org/10.1016/j.jnca.2016.03.014.

[26] V. Raychoudhury, J. Cao, M. Kumar, D. Zhang, Middleware for pervasive computing: a survey, Pervasive Mobile Comput. 9 (2) (2013) 177–200. http://dx.doi.org/10.1016/j.pmcj.2012.08.006. Special Section: Mobile Interactions with the Real World.

[27] 3GPP, TR. 45.820 Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT), Tech. rep., Third Generation Partnership Project, 2015. URL www.3gpp.org/dynareport/45820.htm.

[28] Y.C. Hu, M. Patel, D. Sabella, N. Sprecher, V. Young, ETSI White Paper No. 11. Mobile Edge Computing A key technology towards 5G, White paper, ETSI, 2015. URL http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf.

[29] O. Mkinen, Streaming at the edge: local service concepts utilizing mobile edge computing, in: Next Generation Mobile Applications, Services and Technologies, 2015 9th International Conference on, 2015, pp. 1–6. http://dx.doi.org/10.1109/NGMAST.2015.35.

[30] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, Fog computing and its role in the internet of things, in: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, in: MCC'12, ACM, New York, NY, USA, 2012, pp. 13–16. http://dx.doi.org/10.1145/2342509.2342513. URL http://doi.acm.org/10.1145/2342509.2342513.

[31] 3GPP, TS 36.101 policy and charging control architecture, Tech. rep., Third Generation Partnership Project, 2016. URL www.3gpp.org/dynareport/36101.htm.

[32] 3GPP, TS 36.521 Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) conformance specification; Radio transmission and reception; Part 1: Conformance testing, Tech. rep., Third Generation Partnership Project, 2015. URL www.3gpp.org/dynareport/36521-1.htm.

[33] 3GPP, TS 37.901 User Equipment (UE) application layer data throughput performance, Tech. rep., 3GPP, 2015. URL www.3gpp.org/DynaReport/37901.htm.

[34] 3GPP, TS 23.203 Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception, Tech. rep., Third Generation Partnership Project, 2015. URL www.3gpp.org/dynareport/23203.htm.

[35] 3GPP, TS 23.246 Multimedia Broadcast/Multicast Service (MBMS); Architecture and functional description, Tech. rep., Third Generation Partnership Project, 2015. URL www.3gpp.org/DynaReport/23246.htm.

[36] S.B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, S. Lu, Proportional fair frequency-domain packet scheduling for 3GPP LTE Uplink, in: INFOCOM 2009, IEEE, 2009, pp. 2611–2615. http://dx.doi.org/10.1109/INFCOM.2009.5062197.

[37] R. Kwan, C. Leung, J. Zhang, Proportional fair multiuser scheduling in LTE, IEEE Signal Process. Lett. 16 (6) (2009) 461–464. http://dx.doi.org/10.1109/LSP.2009.2016449.

[38] N. Nikaein, M.K. Marina, S. Manickam, A. Dawson, R. Knopp, C. Bonnet, OpenAirInterface: A Flexible Platform for 5G Research, SIGCOMM Comput. Commun. Rev. 44 (5) (2014) 33–38. http://dx.doi.org/10.1145/2677046.2677053. URL http://doi.acm.org/10.1145/2677046.2677053.

[39] X. Foukas, N. Nikaein, M. Kassem, M. Marina, K. Kontovasilis, FlexRAN: A flexible and programmable platform for software-defined radio access networks, in: International Conference on Emerging Networking EXperiments and Technologies, CoNEXT, 2017.

[40] A. Bhamri, N. Nikaein, F. Kaltenberger, J. Hamalainen, R. Knopp, Three-step iterative scheduler for QoS provisioning to users running multiple services in parallel, in: 2014 IEEE 79th Vehicular Technology Conference, VTC Spring, 2014, pp. 1–5. http://dx.doi.org/10.1109/VTCSpring.2014.7023123.